Supplemental Methods

Cell culture and RNA isolation

LCLs were established from B-lymphocytes, cultured, and exposed to 2 μM activated simvastatin or control buffer as previously described[1-3] in 18 statin exposure batches of ≤12 cell lines each. Simvastatin was provided by Merck Inc. (Whitehouse Station, NJ) and activated as previously described.[2] Lysates from approximately 5 million LCLs were homogenized using QIAshredders (QIAGEN), total RNA was extracted using the PureLink RNA Mini Kit (Life Technologies), and RNA integrity was evaluated using an Agilent Bioanalyzer.

RNA-seq library preparation and sequencing

Libraries were prepared in 3 batches and sequenced in 2 batches. In all cases, the paired control and statin samples from the same cell line were processed in the same batch. For library preparation batches 1 and 3, 500 ng of total RNA (RNA integrity number ≥ 8) was made into indexed, strand-specific, paired-end Illumina sequencing libraries by LabCorp (formerly Covance, Seattle, WA) using the TruSeq RNA Sample Prep Kit v2 and corresponding vendor protocol with a final PCR enrichment of 15 cycles with a few modifications. To increase library insert size, the Elute-Prime-Fragment treatment time was reduced to 4 minutes. To yield final libraries containing only the first strand-synthesized product similar to Parkhomchuk et al.,[4] second strand synthesis was conducted using RNAse H and dUTP in place of dTTP. Also, following the final adapter ligation cleanup, libraries were digested with uracil DNA deglycosylase (to remove the dUTP-containing second-strand product). For library preparation batch 2, the total RNA was first polyA-selected using two rounds of the MicroPoly(A)Purist kit (Ambion), and 100 ng of this polyA-selected RNA was used for the remainder of the protocol described above, starting at the Elute-Prime-Fragment step. Libraries were sequenced to an average depth of 67.1 ± 22.3 million 101 (sequencing and library preparation batch 1 (N=80)) or 100 (sequencing batch 2 (N=70); library preparation batches 2 (N=38) and 3 (N=32)) base pair paired-end reads (33.6 million paired-end fragments) on Illumina HiSeq 2000 machines (**Supplementary Figure S3**).

RNA-seq quality control

The sequence read and base quality of all libraries was checked using FastQC prior to alignment. Suspected PCR duplicates were removed from the alignment files using the Picard MarkDuplicates tool. RNA genotypes of 895 SNPs in exonic locations that were well transcribed in the LCLs with a depth of at least 10 reads at the SNP position and for which we had DNA genotype data from the corresponding individuals were called using allele numbers generated with the SAMtools pileup command.[5] Samples with genotype (X chromosome heterozygosity) and RNA-seq data (e.g. expression levels of *XIST* and *USP9Y*) that conflicted with clinically recorded gender information, samples that were suspected sample mixtures based on abnormal allelic ratios across many SNPs, samples whose RNA genotypes did not match well with DNA genotypes of any CAP participants, samples from participants who were likely Latino based on multidimensional scaling analysis in Plink,[6,7] samples that were 5'->3' bias outliers, and LCLs that did not have paired control and statin RNA-seq data were excluded from further analyses.

Gene expression quantification and normalization

Fragments aligning to Ensembl v67 and EBV[8] genes were counted using HTSeq-count with the reverse strand and intersection-strict options. We excluded genes from further analyses if they

had median expression levels below 5 per 10 million aligned fragments overall or in all subsets based on treatment, ethnicity, and/or gender, leaving 13,931 human genes and 73 EBV genes (14,004 total LCL genes). DESeq2[9] was used to adjust the resulting matrix of gene expression level counts from the control and statin samples for library size and to make the dataset approximately homoscedastic (removing most of the dependence of the variance on the mean values) using a variance stabilizing transformation that took the treatment, ethnicity, sex, and sequencing batch of each sample into account. The changes in gene expression were calculated by subtracting the control from the statin-treated variance stabilized expression levels. These variance stabilized deltas were then adjusted for potential confounders prior to correlation analyses with TG statin response using probabilistic estimation of expression residuals (PEER) [26] with K=3 (**Supplementary Figure S4**) since hidden factors 4+ were almost perfectly correlated with hidden factor 3. Notably, since the control and statin-treated LCLs were matched for many potential confounders (for example, they were prepared in the same batches and should have been genetically identical), far fewer hidden factors were needed to adjust the delta data than would be needed for adjustment of the control (endogenous) gene expression data alone. The seven samples from statin exposure batch 1 were excluded from downstream analyses because they clustered separately from the other 150 samples passing the initial QC steps in a graph of PEER hidden factor 2 versus 1, so PEER was rerun excluding those 7 samples to obtain the final residual values used for RNA-seq statistical analyses.

Imputation quality control details

Starting with genotypes of CAP+PRINCE Caucasian (self-reported white) participants generated on the Illumina HumanHap 300K or 610K-Quad, a custom iSelect panel, and the Cardio-MetaboChip[10] platforms, we excluded markers missing >5% genotypes, deviating from Hardy-Weinberg equilibrium ($p<1x10^{-6}$), mapping to multiple regions of the genome, monomorphic within the CAP subset, or known to have >2 alleles in dbSNP. In addition, we excluded individuals whose genetically determined gender differed from the clinical records and individuals who did not cluster with other individuals of European or Middle Eastern ancestry using multi-dimensional scaling analysis in PLINK[7] as previously described.[6]

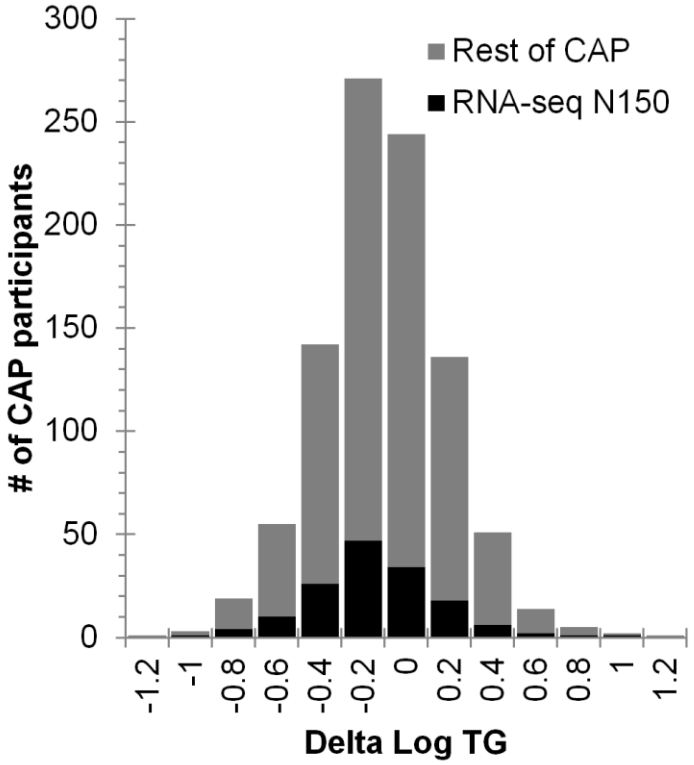

Figure S1. TG statin response distribution in 150 CAP participants with RNA-seq data and the remaining (N=794) CAP participants.

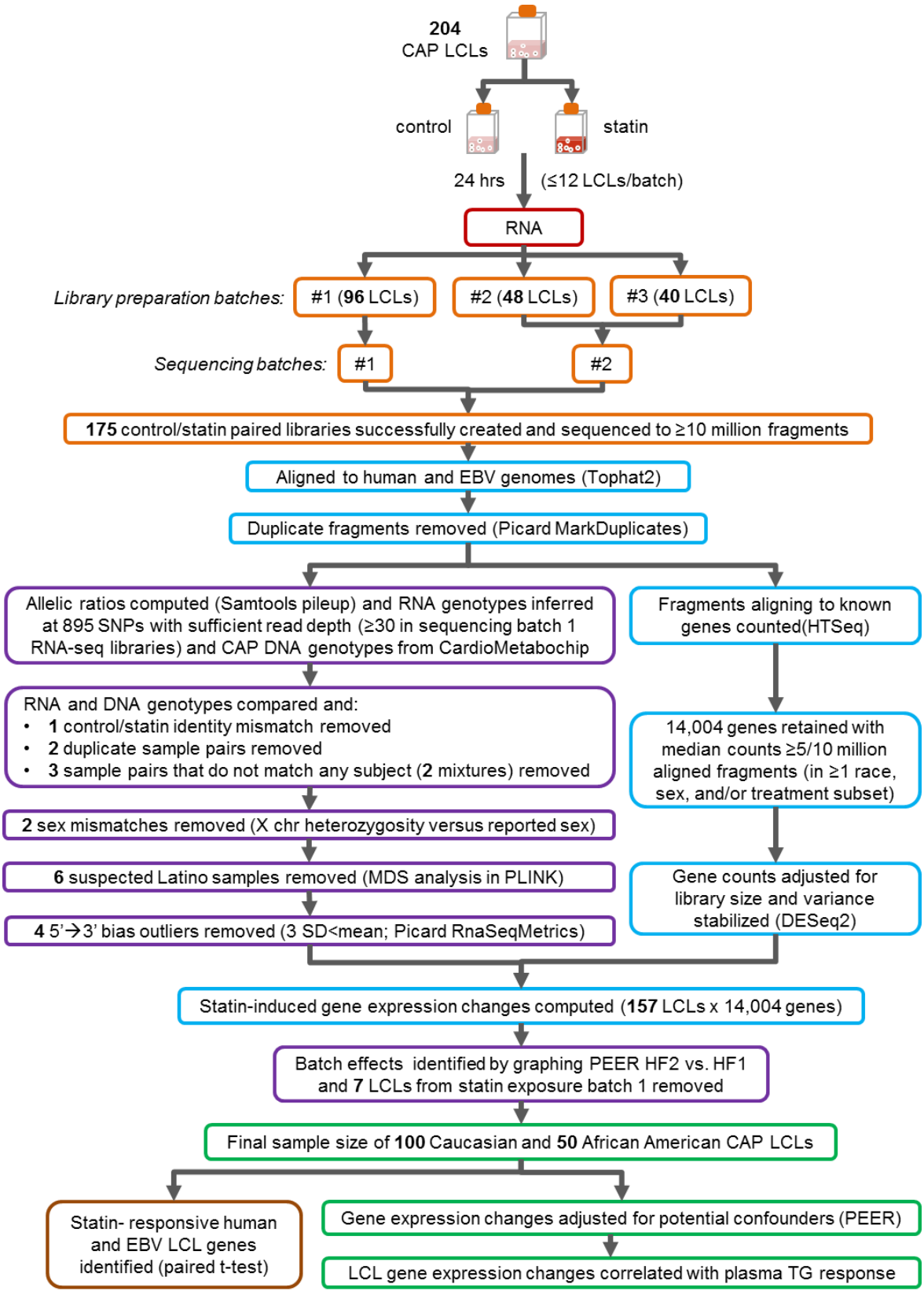

Figure S2. RNA-seq workflow.

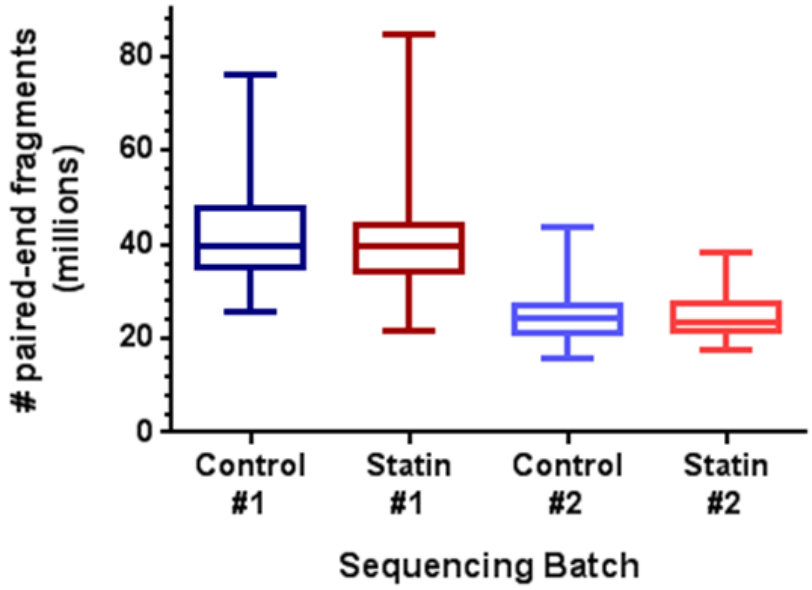

Figure S3. RNA-seq library sequencing depth split by batch and treatment. RNA-seq libraries prepared from 150 control and statin-treated LCLs were sequenced in two batches, the first of which (N=80) was pooled with fewer samples per lane, thus achieving a larger library size than the second batch (N=70) on average. Whiskers are located at minimum and maximum values and boxes show median and quartile values.

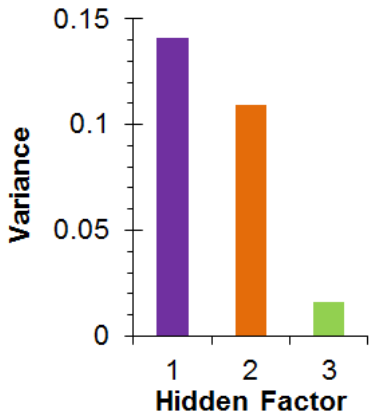

Figure S4. Variance of PEER hidden factors 1-3. Three hidden factors were estimated from the variance stabilized changes in gene expression from the 150 LCLs using Probabilistic Estimation of Expression Residuals (PEER), and the variance of the hidden factors was calculated by taking the inverse of the precision.

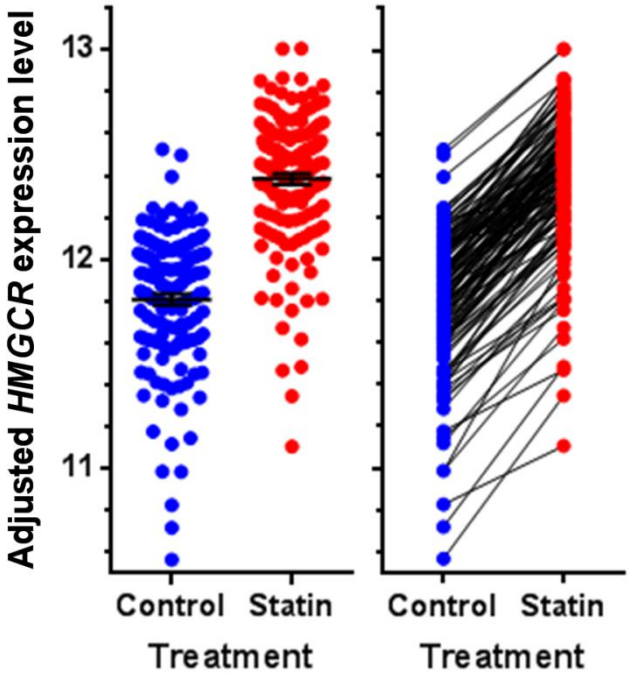

Figure S5. *HMGCR* statin response. Library size adjusted and variance stabilized gene expression levels of *HMGCR*, the direct target of statin treatment, are shown before and after statin treatment in an unpaired (left) and paired (right) fashion for each of the 150 LCLs. Left panel shows mean±SE.

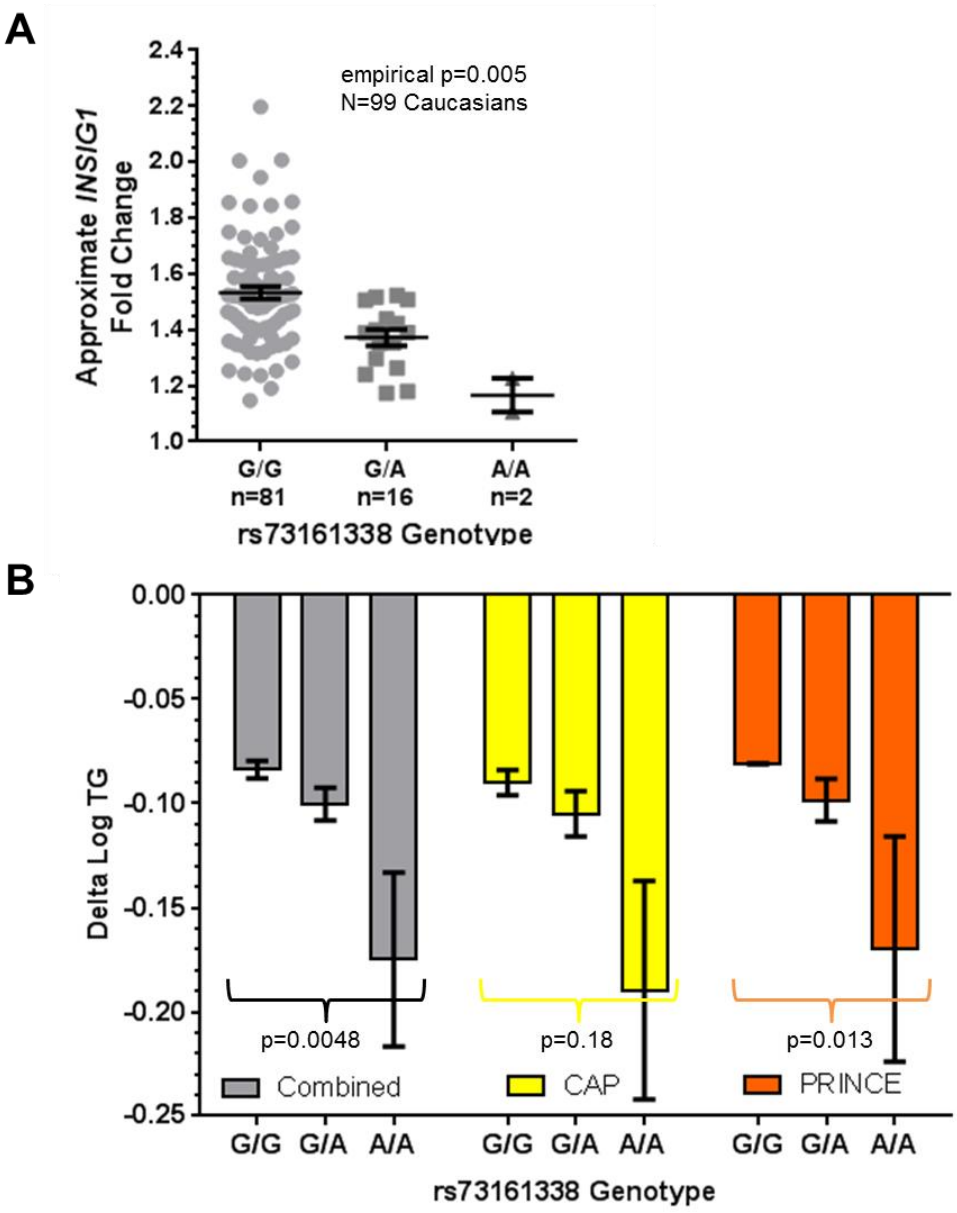

Figure S6. Correlations of *INSIG1* statin response, plasma TG statin response, and rs73161338 genotype. A) *INSIG1* gene expression changes separated by best guess imputed rs73161338 genotype. The unadjusted p-value for the association of the PEER-adjusted variance stabilized deltas versus allelic dosage using an additive allelic model in mach2qtl was 5.4×10^{-5}, and the empirical p-value was calculated from 1,000 simulations across all genetic variants tested in the *INSIG1* genomic region. B) Plasma TG statin response separated by best guess imputed rs73161338 genotype and statin trial. P-values were calculated using an additive allelic model associating allelic dosage with TG response in N=576 CAP and N=1311 PRINCE Caucasian participants. Graphs depict mean±SE.

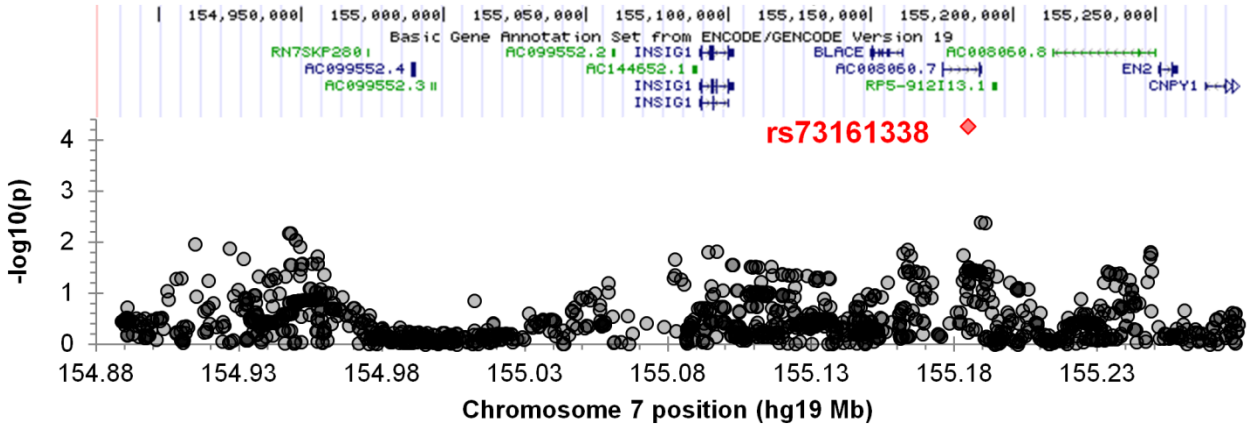

Figure S7. Manhattan plot for local *INSIG1* differential expression QTLs. Common genetic variants within 200 kb of the *INSIG1* transcription start site were tested for association with PEER-adjusted, variance stabilized *INSIG1* gene expression changes using an additive allelic model in mach2qtl in 99 CAP Caucasians with RNA-seq and genotypic data.

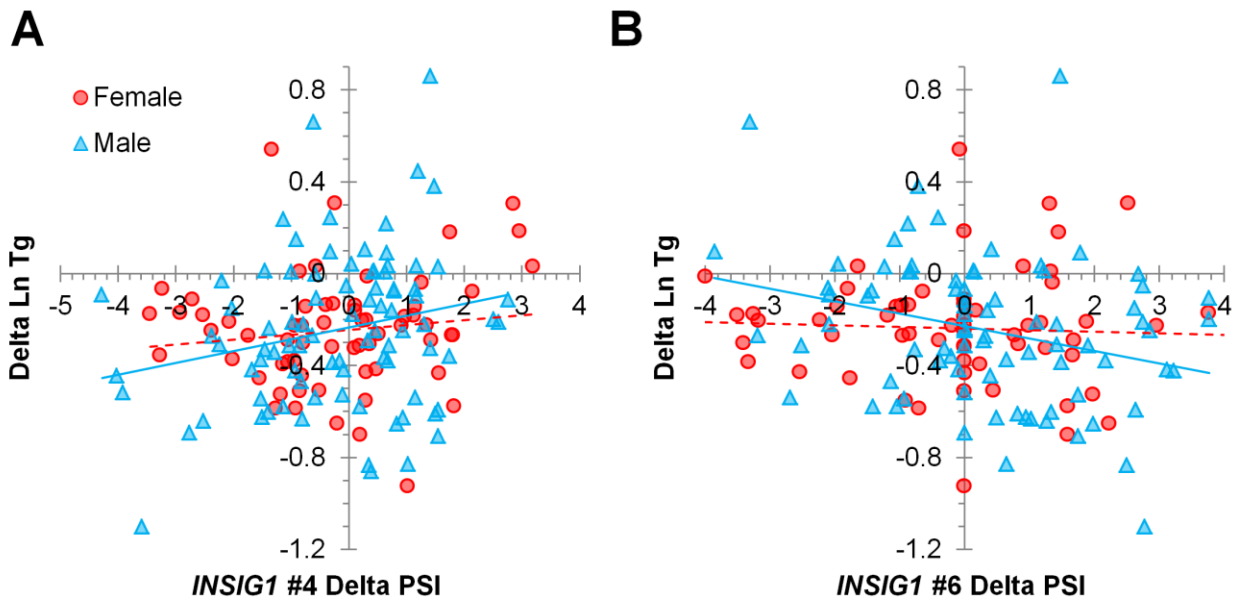

Figure S8. Correlations of changes in *INSIG1* percent-spliced-in values with plasma TG statin response split by sex. After quantifying PSI values for common *INSIG1* splice variants in control- and statin-treated LCLs using JuncBASE, A) statin-induced changes in Event #4 PSI values were tested for correlation with delta log TG using Spearman's correlation in men (N=86, rho=0.19, p=0.09) and women (N=64, rho=0.13, p=0.30) separately, and B) statin-induced changes in Event #6 PSI values were tested for correlation with delta log TG using Spearman's correlation in men (N=81, rho=-0.30, p=0.007) and women (N=61, rho=-0.12, p=0.35) separately.

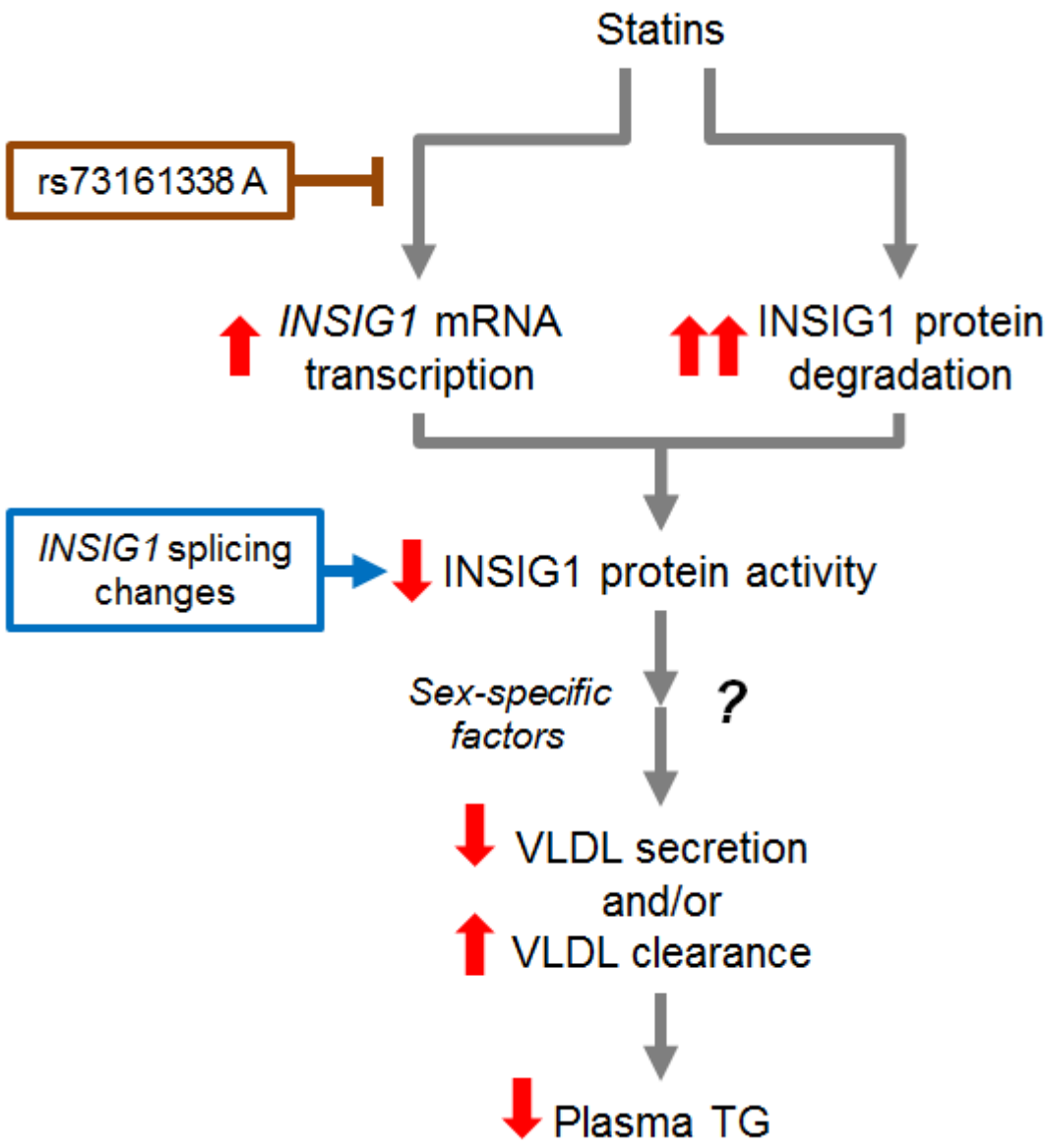

Figure S9. Hypothetical model for *INSIG1* influences on plasma triglyceride changes. Statin treatment causes sterol depletion, which stimulates INSIG1 to dissociate from SCAP in the ER, increasing levels of nuclear SREBFs and increasing transcription of *INSIG1*, with smaller increases in *INSIG1* mRNA levels seen in rs73161338 minor allele carriers. Meanwhile, INSIG1 proteins not bound to SCAP are ubiquitinated and degraded, resulting in a net reduction in INSIG1 protein levels. This ultimately leads to a reduction in circulating TG levels, perhaps through an increase in VLDL clearance and/or reduction in VLDL secretion, a process that is likely influenced by sex-specific factors.

Table S2. Gene Ontology biological processes enriched in statin upregulated genes.

Term	Fold Enrichment	p	Bonferroni-adjusted p
GO:0016126~sterol biosynthetic process	3.70	7.04E-09	2.87E-05
GO:0016192~vesicle-mediated transport	1.55	1.46E-08	5.93E-05
GO:0008610~lipid biosynthetic process	1.71	5.03E-07	2.05E-03
GO:0006694~steroid biosynthetic process	2.61	6.03E-07	2.45E-03
GO:0006695~cholesterol biosynthetic process	3.70	2.09E-06	8.46E-03
GO:0007242~intracellular signaling cascade	1.30	3.01E-06	1.22E-02
GO:0008202~steroid metabolic process	1.90	5.72E-06	2.30E-02

Gene enrichment analysis was conducted in DAVID using the 13,931 human LCL genes expressed above our threshold as the background.

Table S3. Gene Ontology biological processes enriched in statin downregulated genes.

Term	Fold Enrichment	p	Bonferroni-adjusted p
GO:0006396~RNA processing	2.09	2.40E-35	9.47E-32
GO:0022613~ribonucleoprotein complex biogenesis	2.73	2.10E-27	8.28E-24
GO:0022403~cell cycle phase	2.09	3.37E-25	1.33E-21
GO:0034660~ncRNA metabolic process	2.46	4.99E-25	1.97E-21
GO:0000278~mitotic cell cycle	2.06	5.73E-23	2.26E-19
GO:0022402~cell cycle process	1.86	1.20E-22	4.73E-19
GO:0000279~M phase	2.14	1.28E-21	5.06E-18
GO:0007049~cell cycle	1.69	1.99E-21	7.83E-18
GO:0034470~ncRNA processing	2.44	2.87E-20	1.13E-16
GO:0006397~mRNA processing	2.08	5.91E-20	2.33E-16
GO:0042254~ribosome biogenesis	2.78	1.28E-19	5.03E-16
GO:0016071~mRNA metabolic process	1.98	1.92E-19	7.57E-16
GO:0008380~RNA splicing	2.11	3.30E-19	1.30E-15
GO:0000375~RNA splicing, via transesterification reactions	2.53	3.68E-19	1.45E-15
GO:0000377~RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	2.53	3.68E-19	1.45E-15
GO:0000398~nuclear mRNA splicing, via spliceosome	2.53	3.68E-19	1.45E-15
GO:0007067~mitosis	2.28	3.84E-19	1.51E-15
GO:0000280~nuclear division	2.28	3.84E-19	1.51E-15
GO:0006259~DNA metabolic process	1.81	9.27E-19	3.65E-15
GO:0006260~DNA replication	2.39	1.06E-18	4.18E-15
GO:0048285~organelle fission	2.23	1.54E-18	6.08E-15
GO:0000087~M phase of mitotic cell cycle	2.24	2.23E-18	8.80E-15
GO:0016072~rRNA metabolic process	2.73	1.39E-14	5.51E-11
GO:0006364~rRNA processing	2.72	5.50E-14	2.16E-10
GO:0006399~tRNA metabolic process	2.42	1.13E-12	4.45E-09
GO:0051301~cell division	1.87	1.75E-12	6.90E-09
GO:0007059~chromosome segregation	2.75	3.60E-12	1.42E-08
GO:0006261~DNA-dependent DNA replication	2.91	8.22E-11	3.24E-07
GO:0034621~cellular macromolecular complex subunit organization	1.76	1.28E-10	5.02E-07
GO:0043933~macromolecular complex subunit organization	1.49	1.31E-09	5.18E-06
GO:0034622~cellular macromolecular complex assembly	1.77	1.52E-09	5.99E-06
GO:0000070~mitotic sister chromatid segregation	3.39	2.56E-09	1.01E-05
GO:0022618~ribonucleoprotein complex assembly	2.56	3.39E-09	1.34E-05
GO:0015931~nucleobase, nucleoside, nucleotide and nucleic acid transport	2.24	5.51E-09	2.17E-05
GO:0000819~sister chromatid segregation	3.29	6.69E-09	2.64E-05
GO:0006974~response to DNA damage stimulus	1.60	7.80E-09	3.07E-05
GO:0065003~macromolecular complex assembly	1.48	9.96E-09	3.92E-05
GO:0006281~DNA repair	1.68	1.87E-08	7.35E-05
GO:0051329~interphase of mitotic cell cycle	2.21	2.70E-08	1.06E-04

GO:0050657~nucleic acid transport	2.23	3.91E-08	1.54E-04
GO:0051236~establishment of RNA localization	2.23	3.91E-08	1.54E-04
GO:0050658~RNA transport	2.23	3.91E-08	1.54E-04
GO:0006403~RNA localization	2.21	3.99E-08	1.57E-04
GO:0051325~interphase	2.14	8.82E-08	3.47E-04
GO:0000387~spliceosomal snRNP biogenesis	3.35	1.13E-07	4.45E-04
GO:0008033~tRNA processing	2.32	1.67E-07	6.58E-04
GO:0006913~nucleocytoplasmic transport	1.92	2.15E-07	8.45E-04
GO:0051169~nuclear transport	1.91	2.89E-07	0.00113821
GO:0051028~mRNA transport	2.20	4.17E-07	0.001639478
GO:0000059~protein import into nucleus, docking	3.96	8.25E-07	0.003244925
GO:0051327~M phase of meiotic cell cycle	2.23	8.99E-07	0.003533261
GO:0007126~meiosis	2.23	8.99E-07	0.003533261
GO:0051321~meiotic cell cycle	2.17	1.99E-06	0.007818211
GO:0006270~DNA replication initiation	3.92	3.06E-06	0.011983371
GO:0017038~protein import	1.90	4.53E-06	0.017678615
GO:0007005~mitochondrion organization	1.82	7.45E-06	0.028916889
GO:0032508~DNA duplex unwinding	3.52	8.01E-06	0.031070255
GO:0032392~DNA geometric change	3.52	8.01E-06	0.031070255
GO:0051726~regulation of cell cycle	1.51	1.08E-05	0.041507648
GO:0033554~cellular response to stress	1.37	1.30E-05	0.049853114

Gene enrichment analysis was conducted in DAVID using the 13,931 human LCL genes expressed above our threshold as the background.

Table S4. Gene Ontology biological processes enriched in genes with statin-induced changes correlated to TG response.

Term	# genes	p	Bonferroni-adjusted p
GO:0016126~sterol biosynthetic process	6	4.77E-10	6.05E-08
GO:0016125~sterol metabolic process	7	6.41E-10	8.14E-08
GO:0006694~steroid biosynthetic process	6	1.36E-08	1.73E-06
GO:0008202~steroid metabolic process	7	1.90E-08	2.41E-06
GO:0006695~cholesterol biosynthetic process	5	2.04E-08	2.60E-06
GO:0008203~cholesterol metabolic process	6	2.80E-08	3.56E-06
GO:0008610~lipid biosynthetic process	7	7.28E-07	9.24E-05

Gene enrichment analysis was conducted in DAVID using the 13,931 human LCL genes expressed above our threshold as the background. # genes column indicates how many of the 23 genes with statin-induced expression changes significantly correlated with TG response (FDR=15%) were annotated with that GO term.

Table S5. Description of six common *INSIG1* splice variants.

Event #	Description	Junction or Exon Coordinates*
#1	inclusion_junctions	90408-94456, 90376-94456, 94128-94456, 90347-94456, 90095-94456, 94557-99005, 94557-99950
#1	inclusion_exons	94457-94556
#1	exclusion_junctions	94128-99950, 90408-99950, 90408-99005, 90095-99950, 90347-99950, 90376-99950, 94128-99005
#2	inclusion_junctions	94706-95534, 95292-95534, 94557-95534, 95606-99721, 95606-99005, 95606-99950
#2	inclusion_exons	95535-95605
#2	exclusion_junctions	95292-99950, 94557-99005, 94706-99950, 94557-99950, 94557-99721
#3	intron-exon_junctions	87517-87518, 88847-88848, 88871-88872, 89669-89670, 89737-89738
#3	inclusion_junctions	89738-89968
#3	inclusion_exons	87518-88847, 88848-88871, 88872-89669, 89670-89737, 89738-89806
#3	exclusion_junctions	89670-89968, 88872-89968, 89807-89968, 88848-89968, 87518-89968
#4	inclusion_junctions	90408-94456
#4	inclusion_exons	89969-90407
#4	exclusion_junctions	90095-94456, 90347-94456, 90376-94456, 94118-94456, 93443-94456, 94128-94456, 93401-94456
#4	exclusion_exons	93961-94127, 93276-93400
#5	intron-exon_junctions	97079-97080, 97267-97268, 98976-98977, 99005-99006, 99721-99722, 99950-99951
#5	inclusion_junctions	95606-99005
#5	inclusion_exons	96516-97079, 97080-97267, 97268-98976, 98977-99005, 99006-99721, 99722-99950
#5	exclusion_junctions	95606-98976, 95606-96515, 95606-99721, 95606-99950, 95606-97267, 95606-97079
#6	intron-exon_junctions	95605-95606, 99950-99951
#6	exclusion_junctions	95606-99950

Reads crossing the inclusion junctions and falling within the inclusion exon coordinates listed were divided by reads crossing all junctions and exons listed for each splicing event to calculate the percent-spliced-in (PSI) values. *Units of coordinates are the distance (in hg19 bp) from chr7:155.0 Mb.

Table S6. Partial correlations for delta log TG multiple linear regression models

Population	Factor	Squared Partial Correlation (r^2)
Men+Women	*INSIG1* expression change	0.071
Men+Women	Splicing Event #4 change	0.052
Men+Women	Splicing Event #6 change	0.033
Men alone	*INSIG1* expression change	0.172
Men alone	Splicing Event #4 change	0.059
Men alone	Splicing Event #6 change	0.121

Squared partial correlation coefficients were calculated using partial sums of squares.

Supplemental References

1. Medina MW, Gao F, Ruan W, Rotter JI, Krauss RM. Alternative splicing of 3-hydroxy-3-methylglutaryl coenzyme A reductase is associated with plasma low-density lipoprotein cholesterol response to simvastatin. *Circulation*. 2008;118:355-362.
2. Mangravite LM, Medina MW, Cui J, Pressman S, Smith JD, Rieder MJ, et al. Combined influence of LDLR and HMGCR sequence variation on lipid-lowering response to simvastatin. *Arterioscler Thromb Vasc Biol*. 2010;30:1485-1492.
3. Mangravite LM, Engelhardt BE, Medina MW, Smith JD, Brown CD, Chasman DI, et al. A statin-dependent QTL for GATM expression is associated with statin-induced myopathy. *Nature*. 2013;502:377-380.
4. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res*. 2009;37:e123.
5. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078-2079.
6. Theusch E, Medina MW, Rotter JI, Krauss RM. Ancestry and other genetic associations with plasma PCSK9 response to simvastatin. *Pharmacogenet Genomics*. 2014;24:492-500.
7. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-575.
8. Arvey A, Tempera I, Tsai K, Chen HS, Tikhmyanova N, Klichinsky M, et al. An atlas of the Epstein-Barr virus transcriptome and epigenome reveals host-virus regulatory interactions. *Cell Host Microbe*. 2012;12:233-245.
9. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550.
10. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. The metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet*. 2012;8:e1002793.